AI 허위 콘텐츠 대응을 위한 표시제 동향

산업기술리서치센터 초연결신산업팀
박 성 수 (holly_wp@kdb.co.kr)

◆ 생성형 AI의 대중화로 허위 콘텐츠 생성·확산에 따른 부작용이 확대되면서, 중국·미국 등 주요국은 AI 생성물에 워터마크 등 표시를 의무화하는 제도(표시제)* 시행 중
* 콘텐츠가 AI 생성물임을 밝혀 소비자의 진위 파악을 돕기 위해 도입된 제도

◆ '26.1월 국내에서도 표시제를 시행하였으나, 규제 대상이 AI 사업자로 한정되어 이용자·유통자의 표시 삭제·변조 행위 예방에는 한계가 존재함에 따라, 기술·제도적 보완을 통한 실효성 제고 필요

□ AI로 생성된 허위 콘텐츠*가 사회적 혼란을 초래하거나 범죄에 악용되는 사례 증가

* 생성형 AI를 활용하여 텍스트·음성·이미지·영상을 정교하게 합성한 허위 생성물(딥페이크 등)

○ 생성형 AI 기술의 급속한 발전 및 대중화로 비전문가도 콘텐츠를 손쉽게 제작할 수 있게 되면서 AI 생성 콘텐츠의 비중이 빠르게 증가 중

- '25.4월 신규 생성된 영문 웹 페이지 90만 개 중 74.2%*가 AI 생성 콘텐츠를 포함하는 등 콘텐츠 제작 과정에서 AI 활용이 기본으로 정착되는 추세

* Ahrefs('25.5월), "74% of New Webpages Include AI Contents"

○ 생성형 AI를 악용한 허위 콘텐츠도 동반 증가*하면서, 개인 명예훼손, 정치·외교·안보 영역의 혼란, 금융 범죄의 고도화, 행정력 남용 등 부작용 확대

* 국내 딥페이크 범죄 관련 경찰 신고 건수: '21년(156건) → '24년(964건), 6배 증가

- (사례1) '23년 美 대선 당시, 트럼프 대통령이 경찰에 연행되는 허위 콘텐츠 유포로 유권자의 혼란을 초래
- (사례2) '26.4월 동물원 늑대 탈출 사건*은 허위 콘텐츠가 공공 대응 체계 및 행정력 배분을 교란할 수 있음을 보여준 사례

* AI로 합성한 늑대 목격 사진이 실제 제보로 이어지며 재난 문자 발송과 수색 인력이 배치되었던 사건

해외 허위 콘텐츠 사례

자료: 엘리엇 하긴스 엑스(2023)

국내 허위 콘텐츠 사례

자료: 대전소방본부(2026)

□ 생성형 AI의 무분별한 활용 및 허위 정보의 확산을 방지하기 위해 주요국은 AI 생성물 표시 의무 제도 시행 중

○ (중국) AI 생성물에 대한 표시 의무를 체계화하고 생성부터 유통까지 전 과정을 규율화('25.9월 시행)

- AI 생성물의 생산·이용·유통 전 과정을 세분화하여 AI 서비스 제공자에게 명시적 표시*와 암시적 표시**를 모두 이행하는 이중 의무를 부과하고, 이용자·유통자에게 명시적 표시 유지 의무를 부과

* 사용자가 쉽고 명확하게 인지할 수 있는 문자 또는 그래픽 등을 표시

** 파일의 메타데이터에 콘텐츠 속성 정보, 서비스 제공자 명칭, 콘텐츠 번호 등을 표시

○ (미국) 캘리포니아, 뉴욕, 뉴멕시코, 오하이오 등 10여 개 주에서 선거용 딥페이크 광고 표시 의무제를 시행 중이며, 특히 캘리포니아주는 생성형 AI 개발 제작자인 커버드 제공자*를 대상으로 AI 생성물 출처 표시, 무료 탐지 도구 제공, 라이선스 계약 시 라이선시(Licensee)**의 표시 기능 유지를 의무화('26.1월 시행)

* Covered Provider: 월간 방문자·사용자 수 100만 명 이상, 캘리포니아주에서 공개적으로 접근가능한 서비스를 제공하는 자

** 라이선스 등 권리를 빌려서 사용하는 사람이나 기업

- 캘리포니아는 AI 생성물 이용자·유통자의 표시 위조·삭제 가능성을 고려하여 커버드 제공자에게 검증 시스템 구축 의무를 부과한 것이 특징

○ (유럽 연합) 생성형 AI에 의한 생성·조작 여부를 감지할 수 있도록 세계 최초로 법안을 제정하여 AI 서비스 제공자·이용자에게 기계판독이 가능한 표시 의무를 부과('26.8월 시행)

- AI 생성물이 범죄의 탐지·예방 및 수사·기소를 위한 경우와 예술적 저작물 등의 일부를 구성하는 경우 표시 예외 조항을 두어 규정 효율성 제고

이미지 내 표시 예시 | **동영상 내 표시 예시**

자료: 과학기술정보통신부(2026), 인공지능 투명성 확보 가이드라인

□ **국내에서는 AI 기본법*을 통해 AI 사업자를 대상으로 표시 의무제를 시행('26.1월) 중이나, 유통자·이용자의 표시 삭제·변조에 대한 대응책 부족**

*「인공지능 발전과 신뢰 기반 조성 등에 관한 기본법」 이하 AI 기본법

○ AI 기본법은 AI 사업자*가 AI 생성물을 제공하는 경우 생성형 AI 또는 딥페이크 프로그램을 사용한 사실을 이용자**가 명확하게 인식할 수 있도록 표시 의무 부과

* AI를 직접 개발·제공하는 자(AI 개발사업자)와 타 AI 개발사업자가 제공한 AI를 자사 제품·서비스의 기능으로 연계·탑재하여 제공하는 사업자(AI 이용사업자)로 구분

** AI 기술이나 서비스를 제공받는 자로서 단순히 업무나 창작의 도구로서 활용하는 자

- AI 사용이 명백하거나 사업자의 내부 업무 처리 목적 등 사전 고지·표시가 불필요한 상황은 표시 의무를 예외로 하여 사업자의 불필요한 부담 완화
- 예술적 표현물에 해당하는 딥페이크 결과물은 전시·향유를 저해하지 않는 범위 내에서 표시 의무를 부과하여 표현의 자유 보장

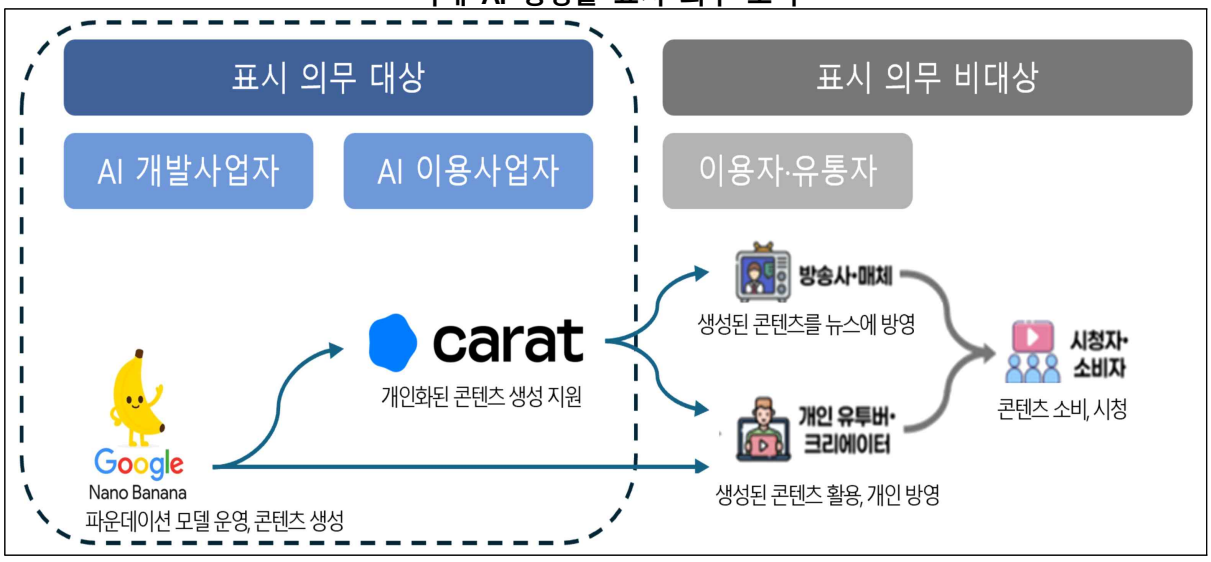
○ 다만, AI 사업자에게만 표시 의무를 부과하고, AI 이용자 및 생성물 유통자에 대한 표시 유지 의무 규정이 부재하여 실효성 한계

- 선거·광고용 AI 생성 허위 콘텐츠에 대한 처벌 규정*이 존재하나 일반적인 이용자·유통자의 부당 사용에 대한 사전 예방에는 한계

* (예시) 공직선거법상 딥페이크 이용 선거운동 제한, 표시광고법상 AI 사용 허위 주장 금지 등

- 표시가 삭제·변조된 AI 생성물이 온라인 플랫폼 등을 통해 재유통되는 경우 진위 여부의 확인 방법이 부재하여 혼란 가중

국내 AI 생성물 표시 의무 도식



자료 : 과기정통부(2026), 인공지능 투명성 확보 가이드라인, 한국산업은행 재구성

□ AI 생성물 표시의 삭제·변조를 방지하고 사후 검증하는 기술·제도적 방안을 마련하여 AI 생성물 표시제의 실효성 제고 필요

○ AI 사업자는 이용자·유통자의 표시 위조·삭제 가능성을 고려하여 제3자가 AI 생성물 여부를 확인할 수 있는 기술적 검증 수단 제공 필요

- AI 생성 이미지, 영상물의 픽셀·주파수 영역 등에 AI 생성 흔적을 삽입하여, AI 생성물 여부를 사후 검증할 수 있는 장치 제공
- 플랫폼 사업자 및 콘텐츠 소비자가 AI 생성물 여부를 직접 확인할 수 있는 검증 프로그램* 및 메타데이터 조회 기능 등 기술적 검증 체계 마련

* (예시) Adobe의 무료 검증 사이트 contentcredentials.org/verify 등

○ AI 생성물의 표시 정보를 임의 삭제·변조하거나, 진위 여부를 오인하게 하는 방식으로 재유통하는 행위를 제한하기 위한 제도적 방안 필요

- SNS 등 콘텐츠 플랫폼 내 의심 게시물에 대한 신고 및 AI 생성 허위 콘텐츠 삭제·차단 등의 절차를 의무화하는 제도 마련

국가별 AI 생성물 표시제 현황

	중 국	미국(California)	E U	한 국
근거 법령	「인공지능 생성·합성 콘텐츠 표시 방법」	「AI 투명성 법」	「인공지능 법」	「AI 기본법」
시행 시기	'25.9월	'26.1월	'26.8월	'26.1월
적용 대상	생산·유통·이용 전 과정의 사업자·이용자	커버드 제공자	AI 서비스 제공자와 이용자	AI 서비스 제공자
적용 범위	전체 AI 생성·합성 콘텐츠	자사 AI 모델이 생성하는 모든 콘텐츠	AI 시스템이 생성·조작한 합성 콘텐츠	생성형 AI를 이용한 제품·서비스의 결과물
특 징	콘텐츠 생산·유통 전 단계를 규율	표시 삭제 검증 시스템 제공	범죄 수사의 용이성, 표현의 자유 보장	AI 사업자만을 대상으로 표시 의무 부과

자료 : 한국콘텐츠진흥원(2025), AI 생성물 표시제 도입에 관한 주요 쟁점과 시사점, 한국산업은행 재구성